Process Performance Evaluation Using Evolutionary Algorithm

*Ahmad S., +Huda S. #Bakir S., *Abdollahian M., *Zeephongsekul P.
*School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia
+School of Information Technology and Mathematical Sciences, University of Ballart, Australia
#School of Management, RMIT University, Melbourne, Australia

Abstract

Nowadays every business is using different quantitative measures and techniques to assess performance of their products/ services. It is well known that different manufacturing processes very often manufacture products with quality characteristics that do not follow normal distribution. In such cases, fitting a known non-normal distribution to these quality characteristics would lead to erroneous results. Furthermore, there is always more than one characteristic Critical to Quality (CTQ) in the process outcomes and very often these quality characteristics are correlated with each other. In this paper, we assess performance of such a bivariate process data which is non-normal as well as correlated. We will use the geometric distance approach to reduce the dimension of the correlated non-normal bivariate data and then fit Burr distribution to the geometric distance variable. The optimal parameters of the fitted Burr distribution are estimated using Evolutionary Algorithm (EA). The results are compared with those using Simulated Annealing (SA) algorithm. The proportion of nonconformance (PNC) for process measurements is then obtained by using the fitted Burr distributions based on the two methods. The results based on both search algorithms are then compared with the exact proportion of nonconformance of the data. Finally, a case study using real data is presented.

Key words: Critical to quality, Burr distribution, Geometric Distance, Proportion of nonconformance, Evolutionary Algorithm, Simulated Annealing.

1. Introduction

Technology has made it relatively easier to record real time data to keep track of process performances. This real time performance tracking record helps managers to improve their business processes and identify which process outcomes have high quality variation when compared with their respective specification spread. If product output falls outside the given specification spread, then, quantitative analysis of critical to quality characteristics is used to identify the proportion of parts being produced outside the customer specifications. Also this analysis provides an opportunity to prevent further production of unacceptable process outcomes.

Traditionally, process performance is estimated using

$$C_p = \frac{USL - LSL}{6\sigma} \qquad (1)$$

where *USL* and *LSL* are upper and lower engineering specification limits respectively. It is a fact that production processes often produce non-normal quality characteristics. Furthermore, there are always more than one quality characteristics of interest in process outcomes and very often they are correlated with each other. Many examples of multivariate non-normal quality data are cited in the quality control literature. This poses the need of multivariate process capability analysis.

In recent years, a multivariate process performance analysis became an interesting research area for many researchers. According to literature review, multivariate process performance, in general, can be obtained from the ratio of a tolerance region to a process region, the probability of nonconforming product, loss function approach and vector representation methods. Another approach has been proposed in the research literature is using computation of higher dimensions except Geometric distance approach and principle component analysis method [12].

Some constraints have been identified in the research literature with regard to the above mentioned multivariate capability analysis such as [11], normality assumption on multivariate data is usually required, confidence intervals of the multivariate capability indices are difficult to derive. Also the computation for high dimension (usually more than three quality variables) characteristics is not easy to perform. Usually, multivariate data are reduced to univariate data before further analysis can be performed; as examples of this, we refer to the geometric distance approach and the principal component analysis method proposed by Wang and Hubele [12], Wang and Du [19].

From above discussions it is clearly evident that application of conventional methods is limited. In order to deal with non-normal multivariate and correlated quality characteristics data, there is an opportunity for researchers to explore suitable performance analysis methods that can address the complex situation of multivariate non-normal data.

In this paper, we employ Wang' geometric distance approach [4] to reduce higher dimensionality of bivariate data and propose to fit just one distribution namely the "Burr distribution" to the geometric distance variable instead of the traditional practice adopted and cited in statistical literature, i.e. fitting different distributions to the geometric data [4, 18]. Furthermore, systematic random search methods called evolutionary and simulated annealing algorithms are introduced to estimate parameters of the fitted Burr distribution. The main objective of quantitative analysis is to help quality practitioners to decide whether to accept or reject the process outcomes based on conformance or nonconformance to customer specifications [12]. Keeping in view this fundamental objective, the efficacy of the proposed method is assessed by using the proportion of nonconformance (PNC) criterion to summarize the performance of geometric distance variable.

This paper is organized in the following manner. Review of Burr XII distribution will be discussed in Section 2. Burr distribution fitting and parameter estimation using search methods (evolutionary and simulated annealing) is discussed in Section 3. An application example with real data using proposed methodology is presented in Section 3. Finally conclusion along with suggestions for future research is given in Section 4.

2. Geometric Distance (GD) Approach

Geometric Distance approach was proposed by Wang and Hubele [18]. The strength of this method is that it reduces the dimension of the multivariate and bivariate process data and renders them more tractable for statistical analysis. The GD approach utilizes the Euclidean distance (or L2 norm) which is defined as follows: let $X = (x_1, x_2, \ldots\ldots, x_n)$ represent a point from a sample space and let $T = (t_1, t_2, \ldots\ldots, t_n)$ be the corresponding target value. Then the Geometric Distance (GDV) variable is defined by

$$GDV = \sqrt{(X - T)' (X - T)}$$

$$GDV = \sqrt{(x_1 - t_1)^2 + (x_2 - t_2)^2 + \ldots + (x_n - t_n)^2} \qquad (1)$$

A comprehensive study of the distribution of G when the underlying variables have a multivariate normal distribution was undertaken in [18]. When the underlying distribution is non-normal, Wang [4] combined correlated quality characteristics to form GDV and determined the distribution that best fits GDV by using Best-Fit statistical software.

In this paper, instead of using different distribution as practiced in the Best-fit approach [4], we will fit just one distribution, the Burr XII distribution to the geometric distance variable. We propose to use evolutionary and simulated annealing algorithms to estimate the parameters of the fitted Burr distribution.

The Maximum Radial Distance (MRD) [18] which is used as the upper specification of the geometric distance variables is the distance between the target and the perimeter of the tolerance region. One sided specification as proposed by many researchers [8, 15, 26] is used here as a performance yardstick when the distance data does not follow normal distribution. In this case median = 0 and the upper specification limit (USL) is defined by MRD:

$$MRD = \sqrt{(Tol_{X_1})^2 + (Tol_{X_2})^2} \qquad (3)$$

where Tol_{Xi} = Tolerance perimeter(s) of the quality characteristic $_{Xi}$; i=1,2.

The criterion which is used to assess the efficacy of the proposed method is to determine the proportion of non-conformance as proposed by many researchers in the quality literature. Hence, using MRD as upper bound; the estimated proportion of non-conformance is given by:

$$\text{PNC} = 1 - F(MRD) = 1 - \int_0^{MRD} f(x)\,dx \qquad (2)$$

where $f(x)$ is the density function of the Burr XII distribution. In the next section we will review Burr distribution and then describe the fitted Burr distribution using both search algorithms.

3. Fitting Burr to Geometric Variable Using Evolutionary Algorithm

As mentioned in the earlier section of this paper, geometric distance approach is being used to reduce the dimensionality of bivariate data to univariate data and the Burr distribution is fitted to the new geometric variable. Literature review [1] suggests that Burr XII distribution can easily fit to any real data. Probability density function and cumulative distribution function of Burr XII distribution are defined as follows:

$$f(y) = \begin{cases} \dfrac{cky^{c-1}}{(1+y^c)^{k+1}} & \text{if } y \geq 0; c \geq 1; k \geq 1 \\ 0 & \text{if } y < 0 \end{cases} \qquad (3)$$

$$F(y) = \begin{cases} 1 - \dfrac{1}{(1+y^c)^k} & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \qquad (4)$$

In the above equations, c and k represent the skewness and kurtosis coefficients of the Burr distribution.

In this paper we fit Burr function $f(x)$ to geometric variable data. This data set is obtained from a computer industry [7]. In order to fit the appropriate Burr function, we need to estimate the parameters c and k of Burr function. To accomplish this, we will use the method of Maximum Likelihood Estimation (MLE) to estimate these Burr distribution (c and k) parameters. The likelihood function of Burr XII distribution is defined by

$$L(c, k; x_1, \ldots, x_n) = \frac{c^n k^n \prod_{i=1}^{n} (x_i)^{c-1}}{\prod_{i=1}^{n} (1 + x_i^c)^{k+1}} \qquad (5)$$

The corresponding log-likelihood function is defined by

$$\log L = n \log(c) + \log(k) - (1+k) * \cdot$$
$$* \sum_{i=1}^{n} \log(1 + x_i^c) + (c-1) \sum_{i=1}^{n} \log x_i \qquad (6)$$

The first order condition for obtaining optimal c and k values gives rise to the following differential equations:

$$\frac{\partial l}{\partial c} = \frac{n}{c} + \sum_{i=1}^{n} \log x_i - (k+1) * \cdot$$
$$* \sum_{i-1}^{n} \frac{\log x_i \log x_i^c}{1 + x_i^c} = 0 \qquad (7)$$

$$\text{and } \frac{\partial l}{\partial k} = \frac{n}{k} - \sum_{i=1}^{n} \log(1 + x_i^c) = 0 \qquad (8)$$

To determine the MLE estimators of c and k, Evolutionary and Simulated Annealing search methods are used in this paper.

3.1 Evolutionary Algorithm (EA)

Evolutionary algorithm (EA) [16-19] models natural evolution processes. Thus, a typical EA incorporates many of the processes logically similar to the processes of natural evolution including natural selection, genetic operations such as crossover, mutation and fitness evaluation. Table-1 illustrates the basics steps of an EA.

3.1.1 Initial Population of EA

Initial population of the EA comprises a number of chromosomes (solutions) and specifies the starting point of the search. Initial population could be created using random initialization. The main goal of initialization process is to create a population with a good coverage of the search space.

3.1.2 Objective function and fitness evaluation in the EA

The objective function (Log L in equation (6) measures the performance of a chromosome with respect to its parameters and is related to problem under consideration. The value of the objective function for one chromosome is independent of the values of the parameters of other chromosome in the population.

However, the fitness of a chromosome measures its reproductive ability and ability to survive. Unlike the objective function, the fitness of a chromosome is always defined with respect to other chromosomes of the population being assessed. The fitness function transforms the value of objective function into a measure of reproductive ability. The chromosomes are ordered according to the objective function values in descending. The fitness is computed using equation (9).

$$Fitness = 2 - SP + 2 * (SP - 1) * (pos - 1)/(PS - 1) \qquad (9)$$

where SP=Selection pressure= {1.0, 2.0}, Pos=Ranked position is the ordered chromosome in the ordered population, PS=Total chromosome in the population

Table 1: Evolutionary Algorithm (EA)

- Create an initial Population.
- Evaluate the initial population using the objective function and order the population based on objective function in descending order.
- Compute the fitness of the population.
- Build the parent pool (mating population) using selection operator.
- Apply crossover operator to create the offspring population.
- Apply mutation operator to the offspring population.
- Apply replacement strategy to form the new population for next generation.
- Return the best solution from the final population.

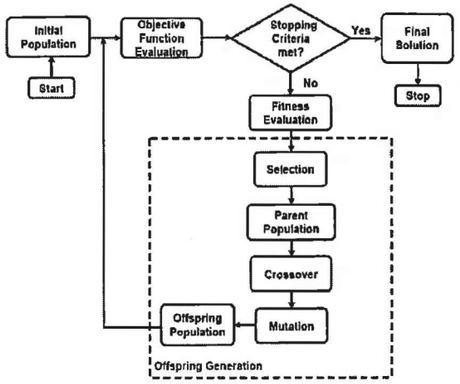

Figure 1: The structure of a simple Evolutionary Algorithm

3.1.3 Selection operator and Parent pool:

Selection operator is used to build a pool of parent chromosomes for reproduction of offspring chromosomes for next generation from the current population. According to Schema theorem [16], reproductive opportunities are allocated to each chromosome in the current population in accordance to proportion of their relativeness. Therefore, chromosome with higher fitness gets higher probability of being selected for reproduction. Thus, the selection operator is implementing a survival-of-the-fittest strategy to build the parent pool. A stochastic Sampling and Replacement selection method have been used here [16, 17].

3.1.4 Crossover operator and Offspring chromosomes

The crossover operator produces offspring chromosomes by exchanging some genetic material between the two parent chromosomes. During crossover, a pair of parent chromosomes (P_1 and P_2) is randomly selected from the parent pool. Then, a crossover operator is applied on (P_1 and P_2), which does some exchange and reordering of information in parent chromosomes and produces two offspring ($OffSpring_1$ and $OffSpring_2$). We have used the following arithmetic crossover [18] operator as follows in equation- (10, 11).

$$OffSpring_1 Var_r = P_1 Var_r + \alpha(P_2 Var_r - P_1 Var_r) \quad (10)$$

$$OffSpring_2 Var_r = P_1 Var_r + (1-\alpha)(P_2 Var_r - P_1 Var_r) \quad (11)$$

Where $OffSpring_1 Var_r$ = r-th variable of $OffSpring_1$, $OffSpring_2 Var_r$ = r-th variable of $OffSpring_2$, $P_1 Var_r$ = r-th variable of P_1, and $P_2 Var_r$ = r-th variable of P_2, with r =1, 2, .. , VT. VT = total number of variables, and $\alpha \in \{0, 1\}$.

3.1.5 Mutation Operator

After crossover, the offspring variables are passed through mutation. Mutation operators are stochastic operator which provides small amount randomness to the offspring variables and maintains a sufficient level of genetic variety in the population. This in turn, re-introduces necessary chromosome features into populations that have been unintentionally lost after several generations have passed. We have used the same mutation operator as in the breeder genetic algorithm [19].

3.1.6 Termination Criteria:

The offspring generation and evaluation is continued up to a pre-determined number of EA generation. Then the best solution is chosen from the best chromosomes from all generations.

4. Experiments with real data set:

In this section, we apply our proposed methodology using real data from Wang [7]. The data set is from a computer manufacturing process with two critical to quality characteristics [7]. It contains a sample of 100 parts that were tested on seven quality characteristics of interest to the manufacturer Data set is given in Table 2. In this paper we only consider two characteristics and geometric distance variable (GDV) given below in Table 3

Table 2: critical to quality (CTQ) characteristics data

CTQ1	CTQ2	CTQ1	CTQ2	CTQ1	CTQ2	CTQ1	CTQ2
0.064	0.059	0.097	0.099	0.092	0.097	0.095	0.094
0.099	0.089	0.095	0.093	0.090	0.094	0.072	0.079
0.098	0.097	0.046	0.115	0.084	0.091	0.100	0.093
0.086	0.084	0.076	0.092	0.068	0.074	0.066	0.067
0.080	0.087	0.088	0.073	0.072	0.068	0.078	0.092
0.080	0.074	0.094	0.083	0.071	0.066	0.073	0.079
0.062	0.053	0.097	0.113	0.055	0.041	0.089	0.090
0.057	0.062	0.068	0.065	0.086	0.098	0.096	0.085
0.065	0.052	0.086	0.087	0.087	0.096	0.090	0.089
0.099	0.080	0.075	0.081	0.095	0.088	0.082	0.089
0.056	0.072	0.106	0.101	0.066	0.062	0.080	0.086
0.104	0.092	0.069	0.085	0.096	0.085	0.105	0.110
0.097	0.087	0.076	0.083	0.095	0.116	0.083	0.080
0.100	0.080	0.088	0.082	0.085	0.086	0.078	0.100
0.079	0.079	0.073	0.031	0.104	0.101	0.098	0.112
0.095	0.077	0.100	0.098	0.093	0.114	0.078	0.080
0.072	0.063	0.107	0.105	0.077	0.089	0.117	0.103
0.081	0.058	0.097	0.101	0.089	0.099	0.072	0.094
0.078	0.079	0.104	0.111	0.093	0.102	0.103	0.096
0.085	0.082	0.086	0.089	0.103	0.106	0.087	0.075
0.067	0.062	0.086	0.084	0.069	0.045	0.102	0.114
0.038	0.060	0.086	0.106	0.093	0.075	0.067	0.107
0.109	0.092	0.096	0.109	0.071	0.088	0.087	0.107
0.047	0.055	0.094	0.090	0.087	0.084	0.065	0.100

0.076	0.064	0.088	0.092	0.091	0.091	0.066	0.067

Figure 2 shows the histogram of the real data. From histogram it is clear that both data sets (*CTQ1 & CTQ2*) are non-normal.

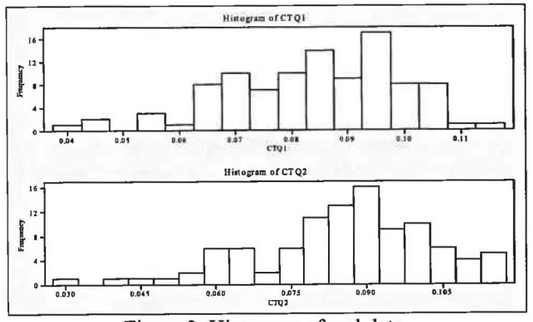

Figure 2: Histogram of real data

Table 3: Geometric distance variable data

GDV	GDV	GDV	GDV	GDV	GDV	GDV
0.013	0.026	0.027	0.021	0.009	0.035	0.013
0.035	0.008	0.051	0.039	0.030	0.010	0.058
0.039	0.017	0.005	0.047	0.052	0.038	0.024
0.021	0.012	0.023	0.031	0.022	0.005	0.042
0.020	0.020	0.012	0.028	0.046	0.023	0.017
0.011	0.008	0.048	0.035	0.049	0.009	0.054
0.019	0.034	0.015	0.031	0.020	0.028	0.037
0.015	0.045	0.014	0.025	0.034	0.030	0.041
0.018	0.027	0.022	0.005	0.039	0.027	0.030
0.031	0.009	0.039	0.003	0.049	0.023	0.005
0.014	0.040	0.041	0.005	0.026	0.019	0.000
0.040	0.034	0.051	0.033	0.023	0.053	0.000
0.031	0.051	0.041	0.032	0.018	0.016	0.000
0.031	0.023	0.053	0.031	0.022	0.031	0.000
0.012	0.018	0.025	0.031	0.030	0.050	0.000

Using a statistical package we found that variables {*CTQ* 1 *and CTQ* 2} are correlated. Correlation and Covariance matrix are given in Table 4 & Table 5.

Table 4: Correlation

Correlations: *CTQ1, CTQ2*

Pearson correlation of CTQ1 and CTQ2 = 0.613

P-Value = 0.000

Table 5: Covariance Matrix

Covariances:	*CTQ1,*	*CTQ2*
	CTQ1	*CTQ2*
CTQ1	0.00022647	
CTQ2	0.00016045	0.00030252

The values for objective functions in the EA generations are presented in the Figure-3. The estimated parameters of the fitted Burr distributions are obtained using Evolutionary Algorithm, Mathematica and Simulated annealing methods and are displayed in Table 6. The proportion of nonconforming data are obtained using equation (2) also presented in table 6.

Table 6: PNC for Geometric Distance Data

Burr parameter estimation method	MRD	c	k	Probability of the product conforming	PNC Value
Simulated Annealing	0.071	1.582	236.52	0.97134	0.0287
Mathematica	0.071	2.116	1531.95	0.99639	0.0036
Evolutionary Algorithm (EA)	0.071	2.135	1620.08	0.99651	0.0035

Using the PNC criterion, Table 6 shows that the PNC obtained by using Mathematica and Evolutionary approach are closer to the actual PNC as compared with PNC obtained using the simulated annealing approach. The actual PNC represents the proportion of data that fall outside their respective specification limits given by the computer manufacturer. In this case study, no data point falls outside the respective specification limits. Keeping in view of the above results in Table 6; Evolutionary Algorithm and Mathematica provide a better estimates of PNC value as compared with the results obtained using Simulated Annealing.

5. Conclusion

This paper introduces a new approach (Evolutionary Algorithm) to find the parameters of a fitted Burr distribution to any type of skewed data. Further, the proportion of nonconforming PNC is deployed to assess the efficacy of the proposed approach in finding the suitable Burr distribution. This is achieved by estimating the global parameters using non-constrained based evolutionary and Mathematica search methods. This approach contrasts with that adopted in [12], where different distributions are fitted to different sets of geometric distance data. It is shown that Evolutionary Algorithm approach has lead to PNC value that is much closer to the exact value. We therefore recommend that the proposed method be applied to other non-normal multivariate data to analyze their performance.

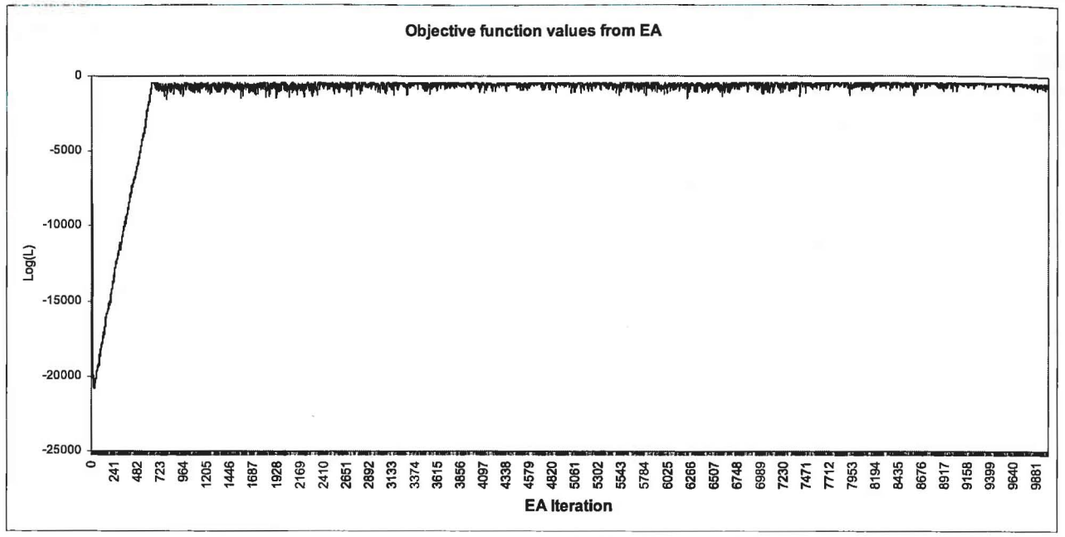

Figure 3. Objective function values form EA generation.

References

1. S. Ahmad, M. Abdollahian, P. Zeephongsekul (2009), Multivariate Non-normal Process Capability Analysis, International Journal of Advance Manufacturing Technology, Int. J Adv Manuf. Technol. DOI (10.1007/s00170-008-1883-9).
2. Bernardo J, Irony T. (1996) "A general multivariate process capability index" The Statistician, 45(4): 487-502.
3. Burr IW (1942) Cumulative frequency distribution. Ann Math Stat 13:215–232
4. Burr IW (1973) Parameters for a general system of distributions to match a grid of á3 and á4. Commun Stat 2:1–21
5. Castagliola P (1996) Evaluation of non-normal process capability indices using Burr's distributions. Qual Eng 8(4):587–593
6. Liu P.H, Chen F.L, (2006), "Process capability analysis of non-normal process data using the Burr XII distribution", Int J Adv manuf Technol 27:975-984.
7. Rodriguez RN (1977). A guide to the Burr type XII distributions. Biometricka, 64:129–134
8. Singpunvalla, N. D. (1992) A Bayesian perspective on Taguchi's approach to quality engineering and tolerance design. IIE Trans., 24, 18-32.
9. Taam, W., Subbaiah, P., Liddy, J. W. (1993). A note on multivariate capability indices. Journal of Applied Statistics, 20(3):339–351.
10. Tang L.C, Than S.E, (1999) "Computing process capability indices for non-normal data: a review and comparative study" Quality and Reliability Engineering International, 15: 339-353.
11. Wierda, S. J. (1993) A multivariate process capability index. In Trans. American Society for Quality Control Quality Congx, pp. 342-348.
12. Wang FK, (2006) Quality evaluation of a manufactured product with multiple characteristics. Qual. Relib. Engng. Int., 22:225-236
13. Wang FK, Du TC. Using principal component analysis in process performance for multivariate data. *Omega* 2000; **28**:185–194.
14. Wang FK, Hubele NF. Quality evaluation using geometric distance approach. *International Journal of Reliability, Quality and Safety Engineering* 1999; **6**:139–153.
15. Wang, Hubele, Lawrence, Miskulin, Shahriari, (2000) "Comparison of Three Multivariate Process Capability Indices," Journal of Quality Technology, 32(3), 263-275.

16. T. Back, Evolutionary Algorithms in Theory and Practice. Oxford University Press, 1996.
17. H. Pholheim, "Genetic and evolutionary algorithm toolbox for use with MATLAB," Dept. Comput. Sci., Univ. Ilmenau, Ilmenau, Germany, 1998. Tech. Rep.
18. Z. Michalewicz and M. Schoenauer, "Evolutionary algorithms for constrained parameter optimization problems," *Evol. Comput.*, vol. 4, no. 1, pp. 1–32, 1996.
19. H. Muhlenbein, "Predictive models for the breeder genetic algorithm continuous parameter optimization," *Evol. Comput.*, vol. 1, no. 1, pp. 25–49, 1993.